Adaptive depletion for improvement of MPEG video compression

Daniel Tan, Svetha Venkatesh, Geoff A.W. West
Department of Computer Science,
Curtin University of Technology,
Bentley, Perth,
Western Australia
email: tand,svetha,geoff@cs.curtin.edu.au

1 Abstract

Traditional data compression algorithms for 2D images work using the information theoretic paradigm, attempting to reduce redundant information by as much as possible. However, through the use of a depletion algorithm that takes advantage of characteristics of the human visual system, images can be displayed using only half or a quarter of the original information with no appreciable loss of quality.

The characteristics of the human visual system that allows the viewer to perceive a higher rate of information than is actually displayed is known as the beta or picket fence effect. It is called the picket fence effect because its effect is noticeable when a person is travelling along a picket fence. Despite the person not having an unimpeded view of the objects behind the fence at any instant, as the person is moving, the objects behind the picket fence are clearly visible. In fact, in most cases the fence is hardly noticeable at all.

The techniques we have developed uses this effect to achieve higher levels of compression than would otherwise be possible. As a fundamental characteristic of the beta effect is the requirement that there is movement of the fence in relation to the object, the beta effect can only be used in image sequences where movement between the depletion pattern and objects within the image can be achieved.

As MPEG is the recognised standard by which image sequences are coded, compatibility with MPEG is essential. We have modified our technique such that it performs in conjunction with MPEG, providing further compression over MPEG.

Keywords: adaptive spatial subsampling, picket fence effect

0-8194-2042-5/96/$6.00

2 Introduction

Spatial subsampling has been widely used as a means to reduce redundant information in the form of fixed spatial subsampling (e.g. MUSE),[32,17] and adaptive spatial subsampling.[22,30,6,26,15,16] Motion compensation is also used with subsampling to achieve higher levels of compression and/or lower the error rate.[24,18,23,32,17,7] While spatial characteristics of the human visual system have been exploited to achieve compression through techniques such as the DCT quantization matrix,[29,27,9,21] vector quantization,[25,31] and subband coding,[8,13,14] little has been done to exploit the temporal characteristics of the human visual system.[10,28] In this paper, we present two techniques which rely on the visual system's abililty to spatially interpolate for moving sequences, and simulate the visual system's ability to integrate over space and time.

Research into the ablility of the human visual system to spatially interpolate was due to the discovery that for moving objects, the human visual system is able to function using significantly less information than previously assumed.[4] An example of this is the beta or picket fence effect, where an observer views a moving object through a picket fence that reveals no more than 10% of the object at any instant, yet is able to see the moving object perfectly.

This phenomenon led to research on the human eye and visual perception system, which resulted in the invention of the Betagraph in 1976. This invention capitalised on the beta phenomenon (hence the name Betagraph) or picket fence effect.[33,3,2,11,5,1] Ross and Sala[1] proposed that instead of relying on large numbers of pixels in order to display graphical images of high resolution, by simulating the picket fence effect, the human visual system could be tricked into perceiving a resolution much higher than is actually being displayed. Instead of using a display fully covered with pixels, they suggested the use of columns of pixels spaced far apart (like the gaps in the fence that are the sole source of information). If the image is then moved across the display at a suitable rate, it was discovered that the human visual system would 'fill in the gaps', giving the viewer the perception that the viewed image was being displayed at full resolution. The viewer thus would not notice that the image was depleted.

In our experiments, images are subsampled by depleting the columns and rows of redundant information. For the purposes of display, these depleted pixels simply have their luminance value changed to 0, so that they appear black. In this way, subjects can judge the quality of the image based on the number of information pixels remaining, while maintaining the correct position of each information pixel with reference with its neighbours.

3 Depletion

The problem with depletion schemes is that the depletion pattern is visible to the observer. There is, therefore a need to mask these effects. The beta effect[1] has been investigated which has two interpretations:

1. A stationary object with moving depletion patterns — picket fence.
2. A moving object with stationary depletion patterns.

It has been shown that both of these depletion methods allow the observer to see the whole scene without noticing the depletion patterns. To investigate the sensitivity of the observer to the different depletion patterns, a number of experiments were performed.

3.1 Stationary Depletion

Initially, there were six stationary depletion schemes investigated as shown in Figure 1. These patterns were chosen as they represented three different types of depletion (horizontal and vertical, horizontal, and vertical), at two different levels of depletion (1:2 and 1:4). The 1:2 depletion schemes investigated were checkerboard (CB), horizontal (HI), and vertical depletion (VI). The pixel maps of the 1:2 depletion patterns are shown in Figure 1(a-c). The 1:4 depletion schemes investigated were horizontal/vertical (VH), horizontal (HI2), and vertical (VI2). The pixel maps of the six schemes are shown in Figure 1(d-f).

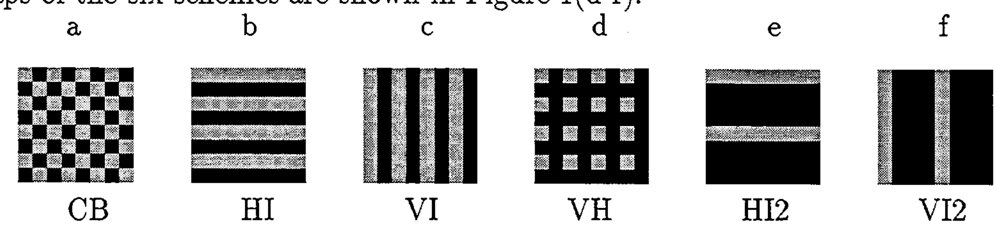

Figure 1: 1:2 and 1:4 Depletion Patterns

The 1:4 horizontal and vertical depletion schemes were rejected as the depletion pattern was very noticeable. An experiment was conducted using the remaining depletion patterns and the original image.

The experiments were run using multiple subjects who assess each of five sequences of movies based on four factors considered important in assessing the quality of the depleted images. The subjects rated each of these factors on a scale ranging from 0 to 9. These four factors were: a) blotchiness, b) edgenoise, c) contrast, and d) resolution.

Blotchiness can best be described as the perceived reduction in the number of greyscale levels in the representation of areas of an image, resulting in large areas of single grey levels, instead of a smooth transition through a number of grey levels. For example, consider an area in an image 10 columns wide, with column 1 having a grey scale value of 0, column 2 having a grey scale value of 1, etc. What the subject would perceive is a smooth transition between columns 1 and 10 from black (grey scale 0) to dark grey (grey scale 9). An example of blotchiness would result in columns 1 to 5 having a grey scale value of 0 and columns 6 to 10 having a grey scale value of 10. Thus, instead of the subject seeing a smooth gradual change in grey scale, they would see two large areas each with a uniform shade of grey.

	VH	VI	HI	CB	OR
Blotchiness	7.3589	7.4211	7.1750	7.7955	6.5897
Std. Dev.	1.3472	1.3878	1.6929	1.3221	1.7429
Edgenoise	7.3333	7.3421	7.3000	7.6136	7.1026
Std. Dev.	1.8543	1.5470	1.5225	1.4975	1.5694
Resolution	7.4872	7.3947	7.4750	7.9318	7.1026
Std. Dev	1.3931	1.2201	1.2808	0.9250	1.4289
Contrast	7.0000	7.0789	7.1000	7.5455	6.4103
Std. Dev.	1.2978	1.2602	1.4815	1.1093	1.6338

Table 1: Average and standard deviation of scores for each depletion method for each of the factors

Edgenoise can best be described as the distraction caused when an edge moves repeatedly from a depleted column or row to a non depleted column or row. When an edge coincides with a non depleted column or row, it merges with the depleted columns or rows on either side of it, forming what appears to be an edge two pixels wider. This illusion ceases when the edge moves and coincides with a depleted column or row. Switching repeatedly between the two possibilities makes the edge appear 'noisy'.

Contrast refers to the grey level resolution of the image, and can be visualised as whether there are sufficient grey scales to clearly represent all the objects within the image. If the subject can clearly distinguish all the detail in the image, whether the portion of the image is light or dark, then the image is considered to have sufficient contrast.

Resolution defines the physical resolution of the image, i.e. whether the pixels are sufficiently small to allow the viewer to resolve small areas of detail. Resolution estimation is usually based on the ability of the subject to see small details within the image clearly. For example, an image in which the subject can clearly make out the pupil of the eye of a person in the image is considered to have a higher level of resolution than another where the whole eye of the person is seen as a grey blob.

In the experiments, the user is first presented with the original luminance corrected frame. The users are then instructed to use this as an example of the ideal case, where it rates 9 in all the factors discussed earlier. The user then views each of the five movies randomly selected, until a predefined number of movies have been viewed. One of the five movies, the luminance normalised original, is the same one presented as the reference by which all the movies in the test are to be rated against. For each of the movies viewed, the user is asked to rate that sequence based on the four previously described factors on a scale of 0 (the worst) to 9 (the best). In the experiments, the scores listed for each of the methods of depletion under each of the factors is determined by taking an average of all the scores given for that method of depletion for that factor by all the subjects.

The results obtained are shown in Table 3.1. In the table, the original movie sequence is labeled OR, with columns labelled CB, HI, VI, and VH, corresponding to the patterns of Figure 1. Note the figures scored by the depletion methods for each factor are similar, but the amounts of depletion are different.

Of interest is the fact that the vertical/horizontal (VH) method of depletion scored similarly to the vertical (VI) and the horizontal (HI) depletion methods despite having half the number of information pixels, and that checkerboard (CB) depletion scored higher than vertical (VI) and horizontal (HI) depletion despite having the same number of pixels of data.

From this experiment, we can draw the conclusion that one of the steps in masking the effect of depletion is to deplete equally along the horizontal and vertical axes.

3.2 Moving Depletion

Having tested stationary depletion with a moving object, the next step was to test moving depletion with a stationary object. The first step of the experiments was to break each image into subimages of 2×2 and 3×3 pixels in size. Within each of these subimages, only one pixel would remain. All the rest would be depleted (and hence have their luminance values changed to 0). Every subimage would have the same corresponding depletion pattern, so that horizontal and vertical distances between pixels would be constant.

The choice of which pixel to remain within each 2×2 and 3×3 subimage was chosen either sequentially or randomly. In the sequential situation, the non-depleted pixels would move sequentially within each subimage, so that a subimage that is 2×2 pixels in size would be completely displayed over four frames, while subimages that are 3×3 pixels in size would take nine frames to completely display each subimage. In the random situation, the non depleted pixel is randomly chosen. As in the sequential situation, it would also take four or nine frames to completely display an image with subimage sizes of 2×2 and 3×3 pixels respectively.

In both seqentially and randomly chosen cases, the results were unsatisfactory. In the sequentially chosen sequences, the cyclic nature of the non depleted pixels were a distraction to the subjects, while in the randomly chosen sequences, the image appeared 'noisy'.

3.3 Adaptive Stationary Depletion

Using the conclusion from the stationary depletion experiment that images have to be depleted equally along both the horizontal and vertical axes to help mask the depletion pattern, we extended the depletion patterns from 1:2 and 1:4 to include 1:8 and 1:16 so as to provide higher levels of depletion. As there are four depletion patterns to choose from, the algorithm has to be adaptive, selecting the depletion pattern depending on the amount of redundancy within the subimage. The pixel maps for these depletion patterns are shown in Figure 2. As such, these represent subsampling of the image by various factors.

Figure 2: 1:2, 1:4, 1:8, and 1:16 Depletion Patterns

These depletion patterns were used simultaneously on each image. To allow adaptive depletion on the whole image, the image is treated as being composed of many subimages. In our case, the subimages are 32×32 pixels in size. The image is first depleted using VH depletion, then interpolated by applying Robert's cross operator[12,19] for interpolation. The standard deviation error between the interpolated subimage and the corresponding subimage within the original image is calculated. If this error is small, the subimage is depleted using a higher level of depletion than if the error is larger. The process by which the images were depleted is shown in Figure 3. The steps denoted by * can be ignored at this stage. Figure 4 represents one frame of the original input image and Figure 5 represents one frame of the adaptively depleted image.

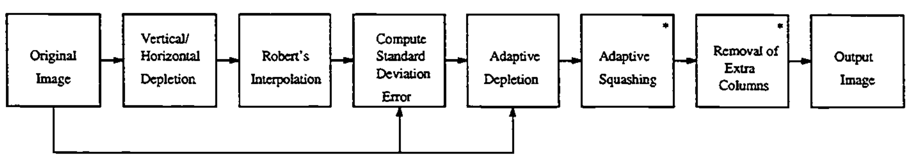

Figure 3: Adaptive Stationary Depletion

When the depleted image is received, interpolation is then performed to reconstruct the original image. The process by which the images are reconstructed is shown in Figure 6. The steps denoted by * can be ignored at this stage. Figure 7 represents one frame of the adaptively repleted image.

The results from this experiment were good, resulting in significant reductions in the number of information pixels with marginal reduction in quality. Subjects reported that the reduction in quality was due to aliasing problems caused by the interpolative reconstruction of the image, earlier described as edgenoise, and the drop in luminance compared with the original image due to the checkerboard depletion pattern.

3.4 Adaptive Moving Depletion

To overcome the problems caused by our interpretation of the beta effect, we have to address the aliasing and luminance reduction problems. To do this, we have to partially simulate

Figure 4: original input image

the psychophysical aspects of the beta effect. Knowing that the beta effect allows the viewer to temporally summate the images seen in such a way that the depleted columns are less noticeable or not noticeable at all, we will simulate this temporal summation by retaining the transmitted information until it is updated by new information. This next experiment differs from the previous one in two ways:

1. The depletion pattern is moving.
2. The displayed image is the sum of the previous images.

As the display is a summation of the previous frames, the luminance problem is overcome. The need for a moving depletion pattern is so that all the pixels within each subimage will at some stage be updated. This also helps us overcome the aliasing problem, as there is no need to interpolate.

As in the previous depletion method, each image is treated as being composed of subimages 32×32 pixels in size. However, instead of having a static depletion pattern, the depletion pattern moves in a systemetic manner over the entire subimage, so that an area depleted using 1:16 depletion would have every pixel displayed over 16 frames, an area depleted using 1:8 depletion would have every pixel displayed over 8 frames, and so on. Because the images displayed at

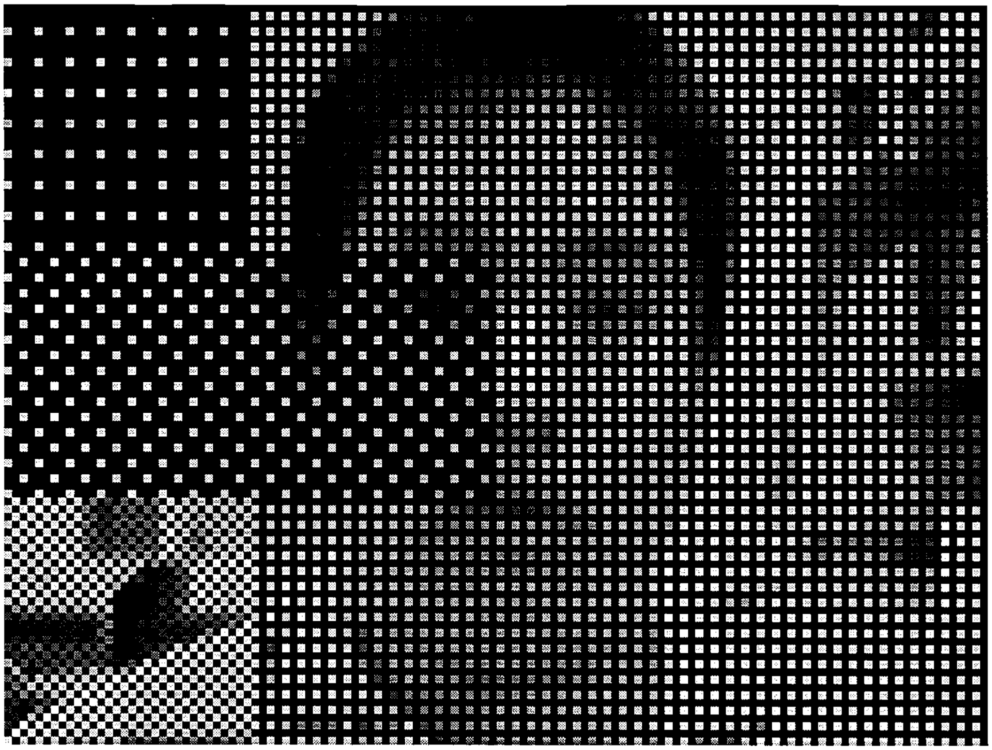

Figure 5: adaptive depletion

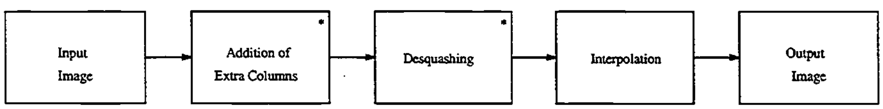

Figure 6: Adaptive Stationary Repletion

the receiving end is a summation of the previous frames, the choice of depletion pattern to use must be related to the difference between the incoming frame and the summation of the previous frames.

When the first frame is transmitted, as there are no previous frames transmitted, the summation of the previous frames (shown in Figure 9 as the summation of pixels) has default values of 0 for every pixel. As the depletion pattern chosen depends on the difference between current frame and the summation of the previous frames, and the difference between each subimage in the copy of the first image and the summation of pixels is usually large, the depletion pattern chosen for the first frame is the checkerboard (CB) pattern. As progressively more images are transmitted and the summation of pixels builds up, incoming images have more similarities to the summation of the previous frames, and thus would be depleted by higher and higher levels of depletion. As a result, only when there are large differences between the incoming image and the previous few images (such as would occur when a scene changes) would lower levels of depletion be used. In this way, in the worse case situation, the summated image would be only two frames

Figure 7: adaptive repletion

behind, as it would take only two successive checkerboard depleted images to build up the new scene. The process by which the images were depleted are represented in Figure 9. The steps denoted by * can be ignored at this stage.

At the receiving end, as the displayed image is simply the summation of the previous frames, processing is minimal. When an image is received, it is simply overlaid on the last output image (if it is the first image, the last output image would have default pixel values of 0), and that summation of pixels is displayed. Figure 8 represents the temporally summated image that is displayed. In this way, during drastic scene changes, two successive checkerboard images would be received, which is all that is required to build up a completely new scene. The summation of pixels works by updating the copy of the output image where there are information pixels in the new image, leaving the depleted pixels as the sum of the previous frames. In the earlier section in which moving depletion was tried, the method failed as the user was able to notice the cyclic manner in which the depletion pattern moved. However, when the output image is the sum of the previous images, this problem is significantly reduced. By using the computer to temporally integrate the images, we are able to avoid the luminance and aliasing problems of the adaptive stationary depletion methods.

The process by which the images are reconstructed is shown in Figure 10. The steps denoted by * can be ignored at this stage.

Figure 8: temporal summation image

4 Interaction with MPEG

Ultimately, image compression schemes must be compatible with MPEG or form part of the MPEG standard. By themselves, the depletion patterns are not compatible with MPEG, as they introduce high frequency noise which means there are many high frequency components in the discrete cosine transforms (DCTs) hence resulting in reducing the compression in MPEG. In our experiments, a checkerboard depleted sequence of images would occupy 77% more memory after it was MPEGed than the original sequence.

To reduce the high frequence noise, the information pixels were 'squashed' into the top left hand corner of each subimage (in the case of the figures shown, the lower left hand corner) as shown in Figure 11. As the DCT within MPEG is based on subimages of 8×8, the 'squashed' images have to fall with those boundaries. For the highest factor depletion of 1:16 to fall within those boundaries, the subimages have to be at least 32×32 pixels in size. With this squashing, every 16th to 32nd column would be black. Therefore further compression can be achieved by removing these extra columns before the image is MPEGed, as shown in Figure 12. This results in 8×8 pixel areas of the image being all black. The whole image could thus be squashed into a corner of the image and only the small image sent. However, this was not required as MPEG would require little information to be transmitted or coded for these regions. Also, the small regions which are occupied by information pixels from the image could vary in size between

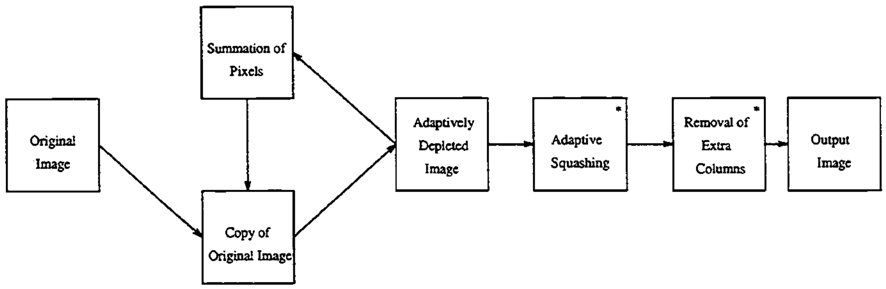

Figure 9: Adaptive Moving Depletion

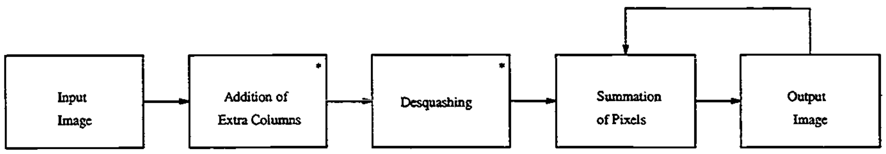

Figure 10: Adaptive Moving Repletion

frames. A table of the compression levels achieved are shown in Table 4.

5 Conclusion

From the preliminary results obtained so far by our experiments, the future for our adaptive subsampling technique looks promising. There are a few areas that still have to be addressed, such as the criteria by which depletion patterns are chosen, but the compression levels achieved over MPEG provide an indication of the potential of this system. The ability for the user to choose the level of the thresholds at which different depletion patterns are used allows the freedom of choice between compression achieved and picture quality.

	Compression	Size(bytes)
Movie, MPEG	100.00	106351
Movie, 1:2 depl., squash., MPEG	81.86	87055
Movie, 1:4 depl., squash., MPEG	78.07	83027
Movie, 1:8 depl., squash., MPEG	62.23	66181
Movie, 1:16 depl., squash., MPEG	39.96	42498
Movie, adapt. depl., squash., MPEG	62.87	66858
Movie, adapt. depl., squash., remove extra cols., MPEG	59.69	63482

Table 2: Compression ratios

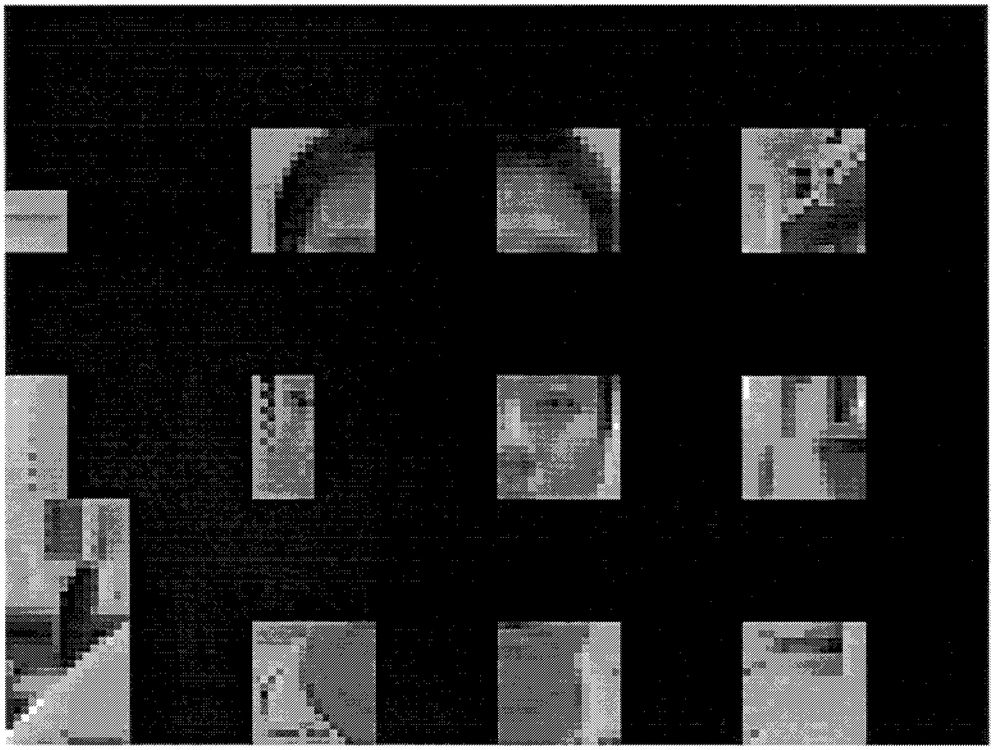

Figure 11: adaptive squashing

6 REFERENCES

[1] Commonwealth Patent No. 21,004/76, 1976.

[2] Anon. Picture devices a hit in U.S. Daily News, January 1976.

[3] Anon. Random scan picture transmission system developed at WA university. Electronic News, March 1976.

[4] Burr D. *On the visibility and appearance of objects in motion.* PhD thesis, Department of experimental psychology and Darwin college, Cambridge, England, 1979.

[5] Burr D. and J. Ross. Visual processing of motion. In *Trends in neuroscience*, pages 304–307, 1986.

[6] Cortelazzo G. and R. Manduchi. On the determination of all sublattices of preassigned index and its application to multidimensional subsampling. *IEEE Transactions on Circuits and Systems for Video Technology*, 3(4):318–320, 1993.

[7] Schroder H. and M. Botteck. Temporal subsampling of spatial highs components in video applications. In H. Yasuda and L. Chiariglione, editors, *Signal processing of HDTV, III*,

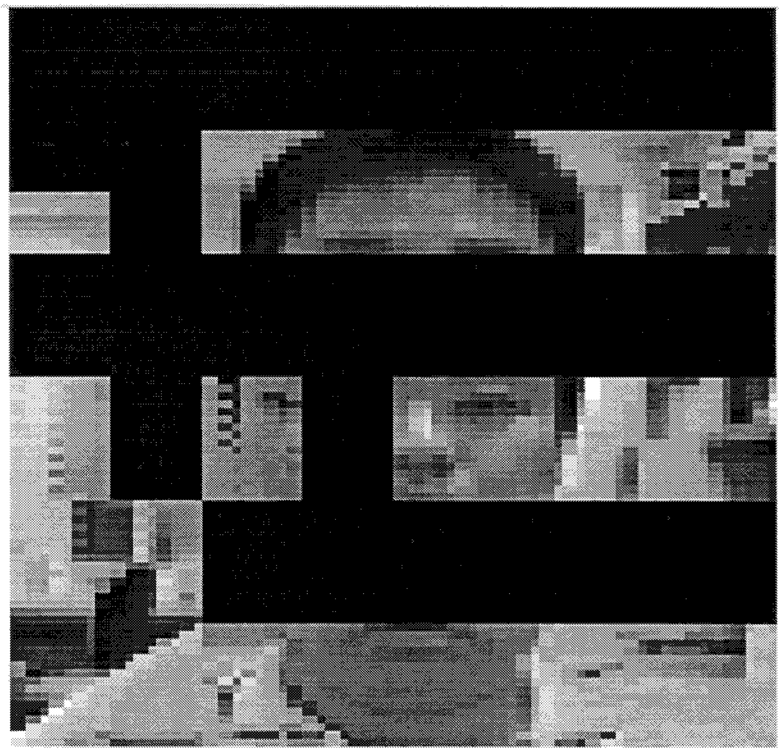

Figure 12: removal of extra columns

pages 379–387, 655 Avenue of the Americas New York New York 10010, 1992. Elsevier Science Publishers.

[8] Schroder H., W. Li, and M. Kunt. High compression of image coding using an adaptive morphological subband decomposition. In *Proceedings of the IEEE*, pages 272–287, 1995.

[9] Yasuda H. A draft iso standard for coded representation of still pictures. In L. Chiariglione, editor, *Signal processing of HDTV*, pages 1–12, 655 Avenue of the Americas New York New York 10010, 1988. Elsevier Science Publishers.

[10] Lee J. and B.W. Dickinson. Temporally adaptive motion interpolation exploiting temporal masking in visual perception. *IEEE Transactions on Image Processing*, 3(5):513–526, 1994.

[11] McIlwraith J. W.A. find boosts graphic transmission. Australian, January 1976.

[12] Staugaard A.C. Jn. *Robotics and AI.* Prentice-Hall, Englewood Cliffs New Jersey 07632, 1 edition, 1987.

[13] Woods J.W. and T. Naveen. Motion compensated multiresolution transmission of HDTV. In *Proceedings of the IEEE global telecommunications conference*, pages 77–79, 1991.

[14] Woods J.W. and S.D. O'Neil. Subband coding of images. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1278–1288, October 1986.

[15] Sawada K. Coding and transmission of high-definition television-technology trends in Japan. *IEEE International Conference on Communications ICC '90*, 3:886–892, 1990.

[16] Ashibe M., K. Mitsuhashi, and S. Tsuruta. A study on adaptive subsampling method for HDTV signal compression. In Chiariglione L., editor, *Signal Processing of HDTV*, pages 145–152, 655 Avenue of the Americas New York New York 10010, 1988. Elsevier Science Publishers.

[17] Itoga M., Y. Mizutani, H. Nakayama, H. Yamamoto, M. Takagi, M. Nakamura, K. Okada, and Y. Izumi. MUSE/NTSC converter. *IEEE transactions on consumer electronics*, 35(3):142–150, 1989.

[18] Tanimoto M., N. Chiba, H. Yasui, and M. Murakami. TAT (Time-Axis Transform) bandwidth compression system of picture signals. *IEEE Transactions on Communications*, 36(3):347–354, 1988.

[19] Fairhurst M.C. *Computer vision for robotic systems.* Prentice Hall International (UK) Ltd, 66 Wood Lane End Hemel Hempstead Hertfordshire HP2 4RG, 1 edition, 1988.

[20] Fleck M.M. Some defects in finite-difference edge finders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):337–344, 1992.

[21] Bousfield P., R. Owens, and J. Ross. JPEG and image measures. In *Proceedings of the 1993 department research conference, department of computer science, University of Western Australia: Perth*, 1993.

[22] Tsang P.M.W. A high quality image compression technique for low cost multimedia applications. *IEEE Transactions on Consumer Electronics*, 41(1):140–149, 1995.

[23] Kishimoto R. and N. Sakurai. A "high-efficiency TCM" bandwidth reduction for high-definition tv. In Chiariglione L., editor, *Signal Processing of HDTV*, pages 129–136, 655 Avenue of the Americas New York New York 10010, 1988. Elsevier Science Publishers.

[24] Belfor R.A.F., M.P.A. Hesp, R.L. Lagendijk, and J. Biemond. Spatially adaptive subsampling of image sequences. *IEEE Transactions on Image Processing*, 3(5):492–500, September 1994.

[25] Gray R.M. Vector quantization. *IEEE ASSP MAGAZINE*, pages 4–29, 1984.

[26] Liu S. and M. Hayes. Variable resolution-adaptive sampling approach to image coding. *IEEE Proceedings of SOUTHEASTCON '91*, 2:1096–1100, 1991.

[27] Pennebaker W.B. and J.L. Mitchell. *JPEG still image compression standard.* Van Nostrand Reinhold, 115 Fifth Avenue New York New York 10003, 1993.

[28] Glen W.E. and K.G. Glen. HDTV compatible transmission system. *SMPTE Journal*, pages 242–246, 1987.

[29] Hobson W.J. and M.P. Gold. Highly efficient still image compression for data broadcasting. *IBC 1990*, pages 223–225, 1990.

[30] Cheung W.N. Adaptive subsampling for image data compression. *IEEE Tencon*, pages 938–941, 1993.

[31] Linde Y., A. Buzo, and R.M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28(1):84–95, January 1980.

[32] Ninomiya Y., Y. Ohtsuka, Y. Izumi, S. Gohshi, and Y. Iwsdate. An HDTV broadcasting system utilizing a bandwidth compression - muse. *IEEE transactions on broadcasting*, 33(4):130–160, 1987.

[33] Kovacs Z. Big future seen for inventions. The West Australian, February 16th 1976.